PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-005000
(43)Date of publication of application : 13.01.1998

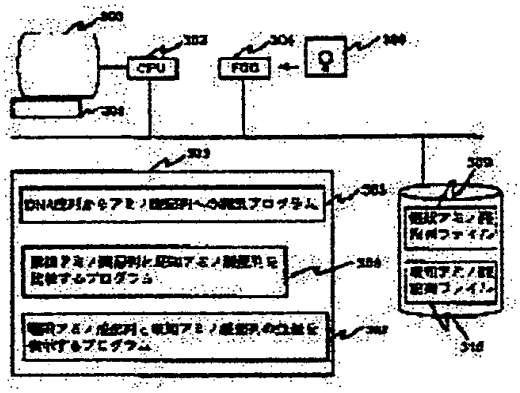
(51)Int.Cl. C12Q 1/68
// C12N 15/09

(21)Application number : 08-167770
(22)Date of filing : 27.06.1996

(71)Applicant : HITACHI LTD
(72)Inventor : KASAHARA NAOKO
NAGAI KEIICHI
HIRAOKA SUSUMU

(54) METHOD FOR COMPARING DNA AMINO ACID SEQUENCE

(57)Abstract:
PROBLEM TO BE SOLVED: To enable the picking of analogous amino acids even in the presence of the insertion or deletion of base units in a DNA sequence by dividing a DNA sequence into base groups of a prescribed length, translating the divided base group into amino acids by shifting the base in a prescribed direction and comparing the translated amino acids with known amino acids.
SOLUTION: A program 305 for translating a DNA sequence to an amino acid sequence, a program 306 for comparing the translated amino acid sequence with a known amino acid sequence by Smith-Waterman method and a program 307 for displaying the translated amino acid sequence together with the known amino acid sequence from the compared result are stored in a main memory 303 of a DNA amino acid sequence comparison apparatus. A DNA sequence is divided into base groups of a prescribed length, the divided base group is translated into amino acids while shifting the base in a prescribed direction and the translated amino acid sequence is displayed together with the known amino acid sequence.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision

of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) 公開特許公報（A）　(11)特許出願公開番号

特開平10－5000

(43)公開日　平成10年(1998) 1月13日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｃ１２Ｑ　1/68	ＺＮＡ	7823－4B	Ｃ１２Ｑ　1/68	ＺＮＡＺ	
// Ｃ１２Ｎ　15/09		9282－4B	Ｃ１２Ｎ　15/00	Ａ	

審査請求　未請求　請求項の数７　ＯＬ　（全 11 頁）

(21)出願番号　特願平8－167770

(22)出願日　平成８年(1996) ６月27日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　笠原　直子
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内

(72)発明者　永井　啓一
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内

(72)発明者　平岡　進
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内

(74)代理人　弁理士　平木　祐輔

(54)【発明の名称】　ＤＮＡアミノ酸配列比較方法

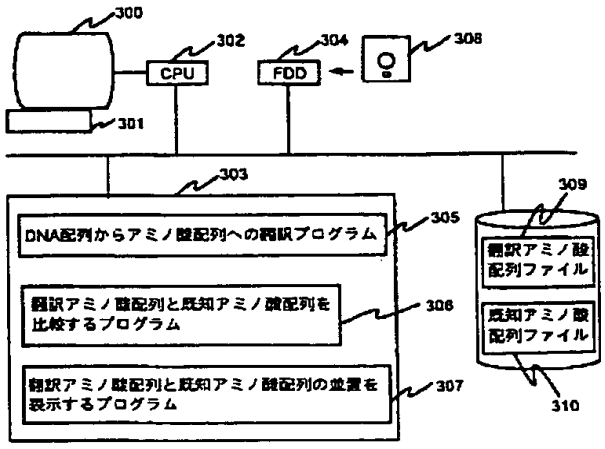
(57)【要約】

【課題】　ＤＮＡ配列に挿入あるいは欠失が存在しても類似している既知アミノ酸配列を高感度に検索できるＤＮＡ配列とアミノ酸配列の直接比較方法を提供する。

【解決手段】　ＤＮＡ配列をアミノ酸への翻訳規則にのっとってアミノ酸配列に翻訳する手段３０５、翻訳アミノ酸配列と既知アミノ酸配列をＤＮＡ配列に存在する挿入あるいは欠失を考慮して比較を行う手段３０６、比較結果に基づいて、翻訳アミノ酸配列と既知アミノ酸配列の並置結果をＤＮＡ配列と共に表示する手段３０７を順次実行し、検索結果を出力する。

【効果】　新たに決定されたＤＮＡ配列に対して配列中に存在する挿入あるいは欠失を考慮して、類似しているアミノ酸配列を高感度に検索を行うことができ、類似している部分を並置結果として表示することで、ＤＮＡ配列の持つ機能を類推することが容易となる。

本発明の配列比較方法を適用する配列比較装置の構成を示す図



【特許請求の範囲】

【請求項１】　塩基の挿入あるいは欠失を含むまたは含まないＤＮＡ配列とアミノ酸配列を直接比較する比較方法において、前記ＤＮＡ配列を所定の長さの塩基群に分割し、分割された塩基群について所定の方向に塩基をずらしてアミノ酸に翻訳し、そのデーターを前記アミノ酸配列と比較して並置を行いその結果を表示することを特徴とするＤＮＡアミノ酸配列比較方法。

【請求項２】　塩基の挿入あるいは欠失を含むまたは含まないＤＮＡ配列とアミノ酸配列を直接比較する比較方法において、前記ＤＮＡ配列を所定の長さの塩基群に分割し、分割された塩基群について５′又は３′末端から１又は２塩基づつずらしてアミノ酸に翻訳し、そのデーターを塩基の挿入あるいは欠失を考慮したあらゆる組み合わせを想定し、最適経路を選ぶ方法により前記アミノ酸配列と比較して並置を行いその結果を表示することを特徴とする請求項１記載のＤＮＡアミノ酸配列比較方法。

【請求項３】　ＤＮＡ配列を５′又は３′末端から所定の方向に塩基を順次シフトして、前記塩基群に分割し、分割されたＤＮＡ配列から翻訳されたアミノ酸配列と比較の対象となるアミノ酸配列の間で、それぞれのアミノ酸について類似度を積算し、その類似度の積算結果が最大となるように、前記ＤＮＡ配列を所定の方向に順次シフトし、翻訳アミノ酸配列を選択することを特徴とする請求項１又は２に記載のＤＮＡアミノ酸配列比較方法。

【請求項４】　ＤＮＡ配列入力手段、前記アミノ酸配列入力手段、ＤＮＡ配列からアミノ酸配列への翻訳手段、該翻訳アミノ酸配列と比較の対象となる前記アミノ酸配列の配列比較手段、該配列比較手段中で類似度を積算する際に参照するスコアテーブル、該翻訳アミノ酸配列と前記アミノ酸配列を並置して前記ＤＮＡ配列とともに表示する手段を有することを特徴とする請求項１乃至３のいずれかの項に記載のＤＮＡアミノ酸配列比較方法。

【請求項５】　請求項３に記載のＤＮＡアミノ酸配列比較方法において、ＤＮＡ配列からアミノ酸配列への翻訳する方法が、ＤＮＡ配列の５′又は３′末端から３文字単位で１文字ずつずらしながら順次翻訳規則にしたがってアミノ酸配列に翻訳する方法であり、ＤＮＡ配列から翻訳されたアミノ酸配列と比較の対象となるアミノ酸配列の配列比較プログラムの類似度を積算する方法が、動的計画法を用いるものであり該動的計画法演算でマトリクスの一方の軸を翻訳アミノ酸配列に他方の軸を比較の対象となる前記アミノ酸配列に対応させた時に翻訳アミノ酸配列のｉ番目の塩基と比較の対象となる前記アミノ酸配列のｊ番目のアミノ酸塩基の対の類似度を積算する際に、前記ＤＮＡ配列中の挿入あるいは欠失が存在する場合と比較の対象となる前記アミノ酸配列中に挿入あるいは欠失が存在した場合を考慮して、（１）ｉ－３番目とｊ－１番目の類似度から積算する場合、（２）ｉ番目とｊ－１番目の類似度から積算する場合、（３）ｉ－３番目とｊ番目の類似度から積算する場合、（４）ｉ－４番目とｊ－１番目の類似度から積算する場合、（５）ｉ－７番目とｊ－２番目の類似度から積算する場合、（６）ｉ－２番目とｊ－５番目の類似度から積算する場合、（７）ｉ－１番目とｊ－５番目の類似度から積算する場合の７種類の経路のうちの少なくとも１つの経路を用いて類似度を積算し、該動的計画法を基に翻訳アミノ酸配列と比較の対象となる前記アミノ酸配列間の類似度の積算値と並置を求めることを特徴とするＤＮＡアミノ酸配列比較方法。

【請求項６】　比較の対象となる前記アミノ酸配列が既知アミノ酸配列であることを特徴とする請求項１乃至５のいずれかの項に記載のＤＮＡアミノ酸配列比較方法。

【請求項７】　前記ＤＮＡ配列が既知ＤＮＡ配列であることを特徴とする請求項１乃至５のいずれかの項に記載のＤＮＡアミノ酸配列比較方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明はＤＮＡアミノ酸配列比較方法に関し、特にＤＮＡ配列とアミノ酸配列を比較しＤＮＡ配列から直接類似アミノ酸配列を検索することに適したＤＮＡアミノ酸配列比較方法である。

【０００２】

【従来の技術】近年、様々な生物のＤＮＡ配列を決定し、その機能を解析する動きが盛んになっている。ＤＮＡ配列はＡ、Ｃ、Ｇ、Ｔの４種類の塩基の並びであり、この塩基の並びの一部がそれぞれ生体中で機能する蛋白質をコードしている。重要な機能を持つ蛋白質は製薬などに利用されることが出来るため、ＤＮＡ配列から直接コードする蛋白質の機能を推定する技術が望まれている。実際に生体中で機能するのは蛋白質配列であるが、ＤＮＡ配列の決定を行なうほうが直接蛋白質配列を決定するよりも技術的に容易であるため、一般的にはＤＮＡ配列を決定する。従って、新たに決定されたＤＮＡ配列の機能を推定するには、機能が分かっている蛋白質と比較し、類似しているか否かを判定することになる。

【０００３】ＤＮＡは３文字の塩基単位（コドン）ごとに２０種類のアミノ酸の一つに翻訳される。ＤＮＡ配列の一部がアミノ酸への翻訳領域であり、その翻訳開始点や終了点の規則は正確には解明されていない。従って、ＤＮＡ配列に塩基レベルの挿入あるいは欠失といった誤りが存在した場合には、ＤＮＡ配列中のアミノ酸への翻訳領域がずれてしまうこともある。

【０００４】また、ＤＮＡ配列はコドン単位でアミノ酸に翻訳されるために、上記の様な誤りがおきた場合には全く異なるアミノ酸として翻訳されてしまうことも考えられる。従来の方法ではこのようなＤＮＡ配列に塩基の挿入あるいは欠失が生じるような誤りに対応した検索を行なっていなかった。従来のＤＮＡ配列とアミノ酸配列

を直接比較する方法として、ＢＬＡＳＴＸ（コロナ社：バイオテクノロジー教科書シリーズ１１「バイオテクノロジーのためのコンピュータ入門」中村春木・中井謙太共著１４１Ｐ－１４３Ｐ、１９９６）がある。この方法では図１で表わされる方法で、ＤＮＡ配列からアミノ酸配列への翻訳を行なう。即ち、まずＤＮＡ配列の端からコドン単位でアミノ酸に翻訳するフレーム（１）、コドンの開始位置をフレーム（１）より１文字ずらしてアミノ酸に翻訳するフレーム（２）、コドンの開始位置をフレーム（１）の開始位置よりも２文字ずらしてアミノ酸に翻訳するフレーム（３）、また該ＤＮＡ配列の相補鎖の反対側の端からコドン単位でアミノ酸への翻訳を開始するフレーム（４）、フレーム（４）の翻訳開始位置から１文字ずらしてアミノ酸への翻訳を開始するフレーム（５）、フレーム（４）の翻訳開始位置から２文字ずらしてアミノ酸への翻訳を開始するフレーム（６）という６つのフレームについてＤＮＡ配列をアミノ酸配列に翻訳する。６種類に翻訳されたアミノ酸配列に対して、それぞれ既知アミノ酸配列と比較を行なう。この方法では、ＤＮＡ配列に塩基単位の挿入あるいは欠失が生じて翻訳フレームが途中でずれてしまった場合に対応していない。例えば、フレーム（１）で翻訳されたアミノ酸配列に非常に類似したアミノ酸配列が存在したとしても、ＤＮＡ配列中に塩基単位で挿入あるいは欠失が存在した場合には、その場所より後はフレームが（２）あるいは（３）に変更される。しかし、従来方法ではそのようなフレームの変更に対応しきれていない。ＢＬＡＳＴＸでは、配列の比較検索を６種類の翻訳フレームを利用し、確率計算を行うことで類似配列を類推する方法を用いている。しかし、この方法では検索もれが生じる可能性がある。また、６種類のフレームのそれぞれについて比較を行なっているために、アミノ酸配列への翻訳が次にどのフレームに変更されているのかが分かりにくいという問題点も生じている。

【０００５】また第２に従来方法として、Ｓｍｉｔｈ－Ｗａｔｅｒｍａｎ法（ＩｄｅｎｔｉｆｉｃａｔｉｏｎｏｆＣｏｍｍｏｎＭｏｌｅｃｕｌａｒＳｕｂｓｅｑｕｅｎｃｅｓ，Ｊ．Ｍｏｌ．Ｂｉｏｌ，（１９８１）．１４７．１９５－１９７，Ｔ．Ｆ．ＳｍｉｔｈａｎｄＭ．Ｓ．Ｗａｔｅｒｍａｎ）がある。この方法は、図２に示したように、比較する２つの配列の文字を１文字ずつ比較して、それぞれに対応したスコアを与え積算し、最終的にスコアが最大となるような検索経路を計算する方法である。この方法は従来ある検索方法の中でもっとも正確な検索方法であるが、配列中の塩基、１文字１文字について比較を行なうために時間がかかる。１組の配列同士のみならず、この場合だと、上記６種類に翻訳されたアミノ酸配列に対してそれぞれ検索を行なわなくてはならないために、さらに検索時間がかかる。また、この方法はアミノ酸単位、あるいはＤＮＡ配列中のコドン単位での挿入あるいは欠失には対応できるものの、ＤＮＡ配列中の塩基単位での挿入あるいは欠失には対応できない。また、この方法でもフレーム間の変更が分かりにくいという問題点も生じる。

【０００６】

【発明が解決しようとしている課題】従来のＤＮＡ配列、アミノ酸配列の比較検索方法では、ＤＮＡ配列の方に塩基単位で挿入あるいは欠失が存在した場合には類似アミノ酸配列を拾ってこないという検索もれをおこす可能性がある。Ｓｍｉｔｈ－Ｗａｔｅｒｍａｎ法では、ＤＮＡ配列を考えうる全てのアミノ酸翻訳フレーム６種類について翻訳し、それぞれに翻訳されたアミノ酸配列を用いて配列比較を行なうために、非常に計算時間がかかる上に異なるフレーム間にまたがった場合にうまく類似部分を特定することが困難である。計算を高速化するために開発された従来方法のひとつであるＢＬＡＳＴＸでは、確率計算を用いて高速化の実現を行なっているためにさらに検索もれを起こす可能性がある。本発明は、そのようなＤＮＡ配列中に存在する塩基単位の挿入あるいは欠失を考慮して、なおかつ検索もれの無いＤＮＡ配列とアミノ酸配列の比較を行なうことが可能とするものである。

【０００７】

【課題を解決するための手段】本発明の特徴は、以下の処理ステップを含むＤＮＡ配列とアミノ酸配列比較方法にある。

〔１〕ＤＮＡ配列を塩基単位の挿入あるいは欠失を考慮してアミノ酸配列に翻訳するステップ。

〔２〕前記ＤＮＡ配列からの翻訳アミノ酸配列と既知アミノ酸配列を挿入あるいは欠失を考慮しつつ配列比較を行なうステップ。

〔３〕比較結果に基づいて、翻訳アミノ酸配列と既知アミノ酸配列の並置結果をＤＮＡ配列とともに表示するステップ。

【０００８】即ち、本発明は塩基の挿入あるいは欠失を含むまたは含まないＤＮＡ配列とアミノ酸配列を直接比較する比較方法において、前記ＤＮＡ配列を所定の長さの塩基群に分割し、分割された塩基群について所定の方向に塩基をずらしてアミノ酸に翻訳し、そのデーターを前記アミノ酸配列と比較して並置を行いその結果を表示することを特徴とするＤＮＡアミノ酸配列比較方法である。

【０００９】更に、本発明は塩基の挿入あるいは欠失を含むまたは含まないＤＮＡ配列とアミノ酸配列を直接比較する比較方法において、前記ＤＮＡ配列を所定の長さの塩基群に分割し、分割された塩基群について５′又は３′末端から１又は２塩基づつずらしてアミノ酸に翻訳し、そのデーターを塩基の挿入あるいは欠失を考慮したあらゆる組み合わせを想定し、最適経路を選ぶ方法により前記アミノ酸配列と比較して並置を行いその結果を表

示することを特徴とするＤＮＡアミノ酸配列比較方法である。

【００１０】更に、本発明はＤＮＡ配列を５′又は３′末端から所定の方向に塩基を順次シフトして、前記塩基群に分割し、前記ＤＮＡ配列から翻訳された該翻訳アミノ酸配列と比較の対象となる前記アミノ酸配列の間で、それぞれのアミノ酸について類似度を積算し、その類似度の積算結果が最大となるように、前記ＤＮＡ配列を所定の方向に順次シフトし、該翻訳アミノ酸配列を選択する事を特徴とする前記ＤＮＡアミノ酸配列比較方法である。

【００１１】更に、本発明はＤＮＡ配列入力手段、前記アミノ酸配列入力手段、ＤＮＡ配列からアミノ酸配列への翻訳手段、該翻訳アミノ酸配列と比較の対象となる前記アミノ酸配列の配列比較手段、該配列比較手段中で類似度を積算する際に参照するスコアテーブル、該翻訳アミノ酸配列と前記アミノ酸配列を並置して前記ＤＮＡ配列とともに表示する手段を有することを特徴とする前記ＤＮＡアミノ酸配列比較方法である。

【００１２】更に、本発明は前記ＤＮＡアミノ酸配列比較方法において、ＤＮＡ配列からアミノ酸配列への翻訳する方法が、ＤＮＡ配列の５′又は３′末端から３文字単位で１文字ずつずらしながら順次翻訳規則にしたがってアミノ酸配列に翻訳する方法であり、ＤＮＡ配列から翻訳されたアミノ酸配列と比較の対象となるアミノ酸配列の配列比較プログラムの類似度を積算する方法が、動的計画法を用いるものであり該動的計画法演算でマトリクスの一方の軸を翻訳アミノ酸配列に他方の軸を比較の対象となる前記アミノ酸配列に対応させた時に翻訳アミノ酸配列のｉ番目の塩基と比較の対象となる前記アミノ酸配列のｊ番目のアミノ酸塩基の対の類似度を積算する際に、前記ＤＮＡ配列中の挿入あるいは欠失が存在する場合と比較の対象となる前記アミノ酸配列中に挿入あるいは欠失が存在した場合を考慮して、（１）ｉ－３番目とｊ－１番目の類似度から積算する場合、（２）ｉ番目とｊ－１番目の類似度から積算する場合、（３）ｉ－３番目とｊ番目の類似度から積算する場合、（４）ｉ－４番目とｊ－１番目の類似度から積算する場合、（５）ｉ－７番目とｊ－２番目の類似度から積算する場合、（６）ｉ－２番目とｊ－５番目の類似度から積算する場合、（７）ｉ－１番目とｊ－５番目の類似度から積算する場合の７種類の経路のうちの少なくとも１つの経路を用いて類似度を積算し、該動的計画法を基に翻訳アミノ酸配列と比較の対象となる前記アミノ酸配列間の類似度の積算値と並置を求めることを特徴とする前記ＤＮＡアミノ酸配列比較方法である。

【００１３】上記比較の対象となるアミノ酸配列としては、例えばアミ酸配列データーベースから選択された既知アミノ酸配列を用いることができる。上記ＤＮＡ配列としては、例えばＤＮＡデーターベースから選択された既知ＤＮＡ配列を用いることができる。

【００１４】

【発明の実施の形態】

【００１５】

【実施例】本発明の第１の実施例について図３を用いて説明する。本実施例はディスプレイ３００、キーボード３０１、中央演算装置ＣＰＵ３０２、主メモリ３０３、フロッピーディスクドライブ３０４から構成される。主メモリ３０３には、ＤＮＡ配列からアミノ酸配列への翻訳プログラム３０５、翻訳アミノ酸配列と既知アミノ酸配列を比較するプログラム３０６、比較した結果から翻訳アミノ酸配列と既知アミノ酸配列の並置を表示するプログラム３０７が格納されている。これらのプログラムはＣＰＵ３０２で実行される。

【００１６】ＤＮＡ配列登録の際には、キーボード３０１から入力されたコマンドにより、ＣＰＵ３０２がフロッピーディスクドライブ３０４に挿入されるフロッピーディスク３０８からＤＮＡ配列を読み取り、ＤＮＡ配列からアミノ酸配列への翻訳プログラム３０５を実行して作成された配列を翻訳アミノ酸配列ファイル３０９として格納する。既知アミノ酸配列は、ＤＮＡ配列と同様にフロッピーディスク３０８から読み込むか、あるいは既に登録されていたアミノ酸配列データベースから読み込み、既知アミノ酸配列ファイル３１０として格納する。

【００１７】配列比較の際には、ＣＰＵ３０２が翻訳アミノ酸配列ファイル３０９と既知アミノ酸配列ファイル３１０から配列を読み込んで、翻訳アミノ酸配列と既知アミノ酸配列を比較するプログラム３０６を実行する。更に、実行結果を用いて、翻訳アミノ酸配列と既知アミノ酸配列の並置を表示するプログラム３０７を実行し、配列比較として出力する。以上が本発明のＤＮＡ配列とアミノ酸配列の直接比較方法を実現するシステムである。

【００１８】以下に、ＤＮＡ配列に塩基単位の挿入あるいは欠失が存在することを考慮して、アミノ酸配列に翻訳するプログラム３０５について説明する。図４に示したように、ＤＮＡは３文字単位のコドン毎に１種類のアミノ酸にコードされる。コドンは４種類のＤＮＡ塩基が３つ組み合って決定されるものなので、６４種類のコドンがあり得る。ところがアミノ酸は２０種類でしかないので、複数のコドンが一つのアミノ酸をコードしていることとなる。このコドンのアミノ酸へのコード規則、すなわち、アミノ酸への翻訳規則を示したのが図４である。

【００１９】つぎに、この翻訳規則を用いてＤＮＡ配列を仮想的にアミノ酸配列に翻訳する。これは、ＤＮＡ配列とアミノ酸配列を直接比較する際に、まず、ＤＮＡ配列を計算上仮想的にアミノ酸配列に翻訳し、そのようにして翻訳されたアミノ酸配列と実際のアミノ酸配列とを比較する方法を用いているからである。従って、図５に

示した方法でＤＮＡ配列をアミノ酸配列に翻訳する。つまりＤＮＡ配列の端からコドンを切り出し該当するアミノ酸に翻訳、次に１文字ずらしてコドンを切り出して同様に該当するアミノ酸に翻訳する。この動作を切り出したコドンの最後の文字が、ＤＮＡ配列の最後の文字になるまでくり返し、最終的にＤＮＡ配列をアミノ酸配列に翻訳する。図５の例では、ＤＮＡ配列がＡＴＧＣＡ・・・ＣＧＡＴなので、まず端から最初のコドンＡＴＧを切り取り対応するアミノ酸Ｍに翻訳する。翻訳アミノ酸配列の１文字目はＭとなる。次にＤＮＡ配列から翻訳アミノ酸配列の２文字目に当たるコドンＴＧＣを、前のコドン開始位置から１文字ずらして切り出しアミノ酸Ｃに翻訳する。更にＤＮＡ配列の１文字ずらした位置からコドンＧＣＡを切り出しアミノ酸Ａに翻訳する。この動作をくり返し、ＤＮＡ配列からアミノ酸配列を翻訳する。図５の例の場合には、翻訳されたアミノ酸配列はＭＣＡ・・・ＲＤとなる。このＤＮＡ配列からアミノ酸配列への仮想の翻訳は、通常のＤＮＡ配列からアミノ酸配列に比較して約３倍量のアミノ酸配列が翻訳されることとなる。この方法を用いることにより、ＤＮＡ配列は１本のアミノ酸配列、相補鎖を考慮してもせいぜい２本のアミノ酸配列に翻訳される。このようにして翻訳されたアミノ酸配列と既知アミノ酸配列を、Ｓｍｉｔｈ－Ｗａｔｅｒｍａｎ法をもとにＤＮＡ塩基単位での挿入あるいは欠失を許容する配列比較方法にて比較する事により、その配列間の類似度を見る事が出来る。

【００２０】以下に、翻訳アミノ酸配列と既知アミノ酸配列間の配列比較のプログラム３０６について詳しく説明する。本発明は、Ｓｍｉｔｈ－Ｗａｔｅｒｍａｎ方法に基づいた配列比較方法を用いている。図６に示されたアミノ酸同士の対に対するスコア表を用いて、翻訳アミノ酸配列と比較対象となっている既知アミノ酸配列の間のスコアを算出し、算出されたスコアに応じてその配列の類似度を類推する方法である。このスコアマトリクスは、アミノ酸の各組の性質の類似度を考慮して、それぞれの組に対してスコアを設定するものである。このマトリクスの値は、どの位の類似度の配列を検索することによって、検索者自身が選択することが可能である。図６は、いくつか実際に使用されているマトリクスの中で、blosum62という名前のマトリクスである。

【００２１】次に、本発明における配列比較方法の中心部分を占めるスコアの算出方法について、図７を用いて説明する。まず、図７に表わされるように翻訳アミノ酸配列と既知アミノ酸配列をマトリクス状に配置する。配置された配列の各要素に対応するアミノ酸同士のスコアは、図６に示されたスコアマトリクスを参照する。配列の要素であるアミノ酸を読み進み、挿入あるいは欠失に対するペナルティを与えながら、スコアを積算していく事でマトリクス内の升を埋めて行く。しかし、この積算時に参照するスコアは図７に示された７つの場合から積算して行く。即ち、図７の（０）の位置のスコアを計算するには、（１）から（７）の位置のスコアに、（０）の位置のアミノ酸対のスコアを図６のマトリクスから参照し加算しまたは適宜に挿入あるいは欠失に対するペナルティスコアを加算して、結果が最大値をとるようなスコアを選択する。この時、（１）から（７）までのどの部分のスコアに加算されたのかを記録しておく。図８に示してあるように、（１）に加算されたスコアが（０）におけるスコアの最大値である時、この場合は翻訳アミノ酸配列と既知アミノ酸配列のアミノ酸が一致していても、不一致であっても対応させる場合である。即ち、翻訳する前のＤＮＡ配列にも既知アミノ酸配列にも挿入あるいは欠失をいれない場合である。（２）に加算されたスコアが最大値となる時は、翻訳アミノ酸配列にアミノ酸１文字の欠失がある場合、即ち、翻訳する前のＤＮＡ配列の該当する部分にコドン単位の欠失が存在する場合である。（３）に加算されたスコアが最大値となる時は、既知アミノ酸配列に１文字の欠失が存在する場合である。（４）に加算されたスコアが最大値となる時は、（０）の部分の翻訳アミノ酸に対応するＤＮＡ配列中のコドンの直前に１塩基の挿入がある場合である。（５）に加算されたスコアが最大値となる時は、（０）の部分の翻訳アミノ酸の直前の翻訳アミノ酸に対応するＤＮＡ配列中のコドンの中に塩基が挿入している場合である。（６）に加算されたスコアが最大値となる時は、（０）の部分の翻訳アミノ酸の直前の翻訳アミノ酸に対応するＤＮＡ配列中のコドン中の塩基が欠失している場合である。（７）に加算されたスコアが最大値となる時は、（０）の部分の翻訳アミノ酸の直前の翻訳アミノ酸に対応するＤＮＡ配列中のコドン単位の挿入とそのコドン中の塩基が欠失している場合である。以上の７つの場合を考慮してスコアの計算を行なう。実際のスコア算出方法に関しては、次に示す。翻訳アミノ酸配列のi番目と既知アミノ酸配列j番目のスコアs(ij)を求める式s(i, j)＝max[score0, score1, score2, score3, score4, score5, score6, score6]

score 0＝score(i, j)＋s(i-1, j-3)
score 1＝s(i-1, j)-4 or s(i-1, j)-12
score 2＝s(i, j-3)-4 or s(i, j-3)-12
score 3＝s(i-1, j-4)＋score(i, j)-12
score 4＝s(i-2, j-7)＋score(i, j)-12
score 5＝s(i-2, j-5)＋score(i, j)-12
score 6＝s(i-1, j-7)＋score(i, j)-24
score 7＝s(i-1, j-5)＋score(i, j)-24

score(i, j)：i番目の塩基とj番目の塩基対に与えられる類似度の指標

式中の減算されている数字(例 -4, -12, -24)は挿入・欠失またはその延長に対して与えられるペナルティスコアである。

【００２２】上記で説明した方法に基づき、翻訳アミノ

酸と既知アミノ酸配列とのスコアを算出する。マトリクスの端でスコアが最大値をとる升を選択し、その最大スコアを配列間の類似スコアとし、配列比較の結果を示す指標とする。このスコアが大きいほどより類似しているＤＮＡ配列とアミノ酸配列であると言う事が出来る。この結果に基づき、翻訳アミノ酸配列と既知アミノ酸配列の並置を表示する。

【００２３】以下に、翻訳アミノ酸配列と既知アミノ酸配列の並置の表示プログラム３０７について、図８を用いて説明する。上記で説明したように、図７で示された（１）から（７）のどの部分のスコアを参照してスコアが算出されたかによって、挿入あるいは欠失の存在する位置が変化する。翻訳アミノ酸配列と既知アミノ酸配列がマトリクス状に配置され、各アミノ酸に対応するスコアが算出された後、配列の一番端を示すマトリクスの行および列上でスコアが最大値をとる位置から、その部分のスコアが図７の（１）から（７）のどの場所から計算されたかをスコアが０になるまでたどって行く。それぞれ（１）から（７）のどの場合でも、図８に対応する並置例をつなげていく事で、最終的に翻訳アミノ酸配列と既知アミノ酸配列の並置を求め、表示する。また、アミノ酸は１文字表記の他に３文字表記も一般的になされているので、翻訳アミノ酸配列と既知アミノ酸配列とを１文字表記で表現するのではなく３文字表記で表示する事で、翻訳アミノ酸配列とＤＮＡ配列をならべて表示する事が出来る。この時、ＤＮＡ配列には図７および図８の規則に従い、挿入あるいは欠失を示す記号を該当箇所に代入する事で、既知アミノ酸配列とＤＮＡ配列との比較をより分かりやすく表示する事が可能である。

【００２４】以下、本発明による比較手順を図９に従って、実際のＤＮＡ配列とアミノ酸配列を用いて説明する。図９中の９０１に示したように、ＤＮＡ配列をａｇｃｔｔｇｃｃａａｃｔとする。図５中で説明した手順に従い、すなわち、ＤＮＡ配列の片方の端から１文字づつずらしながらコドン単位でアミノ酸に翻訳する。コドンがアミノ酸に翻訳される規則は図４に示されている。この規則を用いて上記ａｇｃｔｔｇｃｃａａｃｔというＤＮＡ配列は、図９の９０２に示したようにまず１番端のコドンａｇｃがアミノ酸Ｓｅｒ（１文字表記ではＳ）に翻訳され、次にｇｃｔがアミノ酸Ａｌａ（Ａ）に翻訳される。このように１文字づつずらしながらコドン単位でアミノ酸に翻訳していくという操作を繰り返し、上記ＤＮＡ配列は図９中９０３のアミノ酸配列ＳＡＬＬＣＡＰＱＮＴに翻訳される。この９０３のアミノ酸配列を比較対象となるアミノ酸配列と区別するために翻訳アミノ酸配列と呼ぶことにする。次にこの様にして作成された翻訳アミノ酸配列と、データベース中などの既知のアミノ酸配列との比較方法を説明する。翻訳アミノ酸配列９０３を既知アミノ酸配列９０４と比較する場合を例にとって説明する。図９中９０４の既知アミノ酸配列ＳＡＲＡＰＱＲＤＴと９０３の翻訳アミノ酸配列ＳＡＬＬＣＡＰＱＮＴを比較する場合には以下の手順に従う。まず、９０５に示すように翻訳アミノ酸配列を垂直方向、既知アミノ酸配列を水平方向に配置したマトリクスを考える。基本的な配列比較の方法は、この様にして作成されたマトリクス内の全てのマスにおける類似度の基準となるスコアを算出し、その最大スコアによって、類似しているかしていないかの判別を行う。スコアの算出方法を説明する。図６に示されたようにそれぞれのアミノ酸対には、類似度の指標であるスコアが与えられる。このスコア体系は、求める進化上の距離に応じて選択することができるが、ここでは図６に示したスコアマトリクスＢｌｏｓｕｍ６２を用いる。マトリスク上の各マスにおけるスコアは図７に示したように基本的には既に計算して求められた７つのマスのスコアから算出し、その最大値を選択することによって、該当するマスにおけるスコアを計算する。まず、一番上の行のスコアを計算する。この行は、翻訳アミノ酸配列の一番最初のアミノ酸であるＳと、水平方向に配置された既知のアミノ酸配列９０４のＳＡＲＡＰＱＲＤＴとの間のスコアを算出する。スコアは図に示された７つのマスのスコアから算出されるが、この行のように７つのマスのスコアがまだ計算されていない場合には、スコアの初期値は０として計算を行う。まず、２つの比較する配列９０３と９０４の一番最初のアミノ酸ＳとＳの対に与えられる値は、図６のスコアマトリクスを参照して４であることが分かる。従って、この４という値を図７に示された７つのマスのスコアに加算して、それぞれに得られた値のうち最大値をスコアとする。そのため、図９中の９０５のマトリクスの１行目の最初のマス９０６のスコアは４となる。次に、翻訳アミノ酸配列９０３中の１番目のアミノ酸Ｓと既知のアミノ酸配列９０４中の２番目のアミノ酸Ａとの比較スコア、すなわち、９０５のマトリクス内の１行２列目のマス９０７に該当するスコアを算出する。このマス９０７のスコアは図に示された７つのマスのスコアに、アミノ酸ＳとＡの対に与えられる値１を加算し、その最大値を選択することで求める。ここで、図７に示された７つのマスのうち、（２）に対応するマス９０６以外はスコアが求められていない。従って、このマスのスコアは（２）に対応するマス９０６、すなわち翻訳アミノ酸配列９０３の１番目のアミノ酸Ｓと既知のアミノ酸配列９０４の１番目のアミノ酸Ｓとの比較で算出されたスコア４に、アミノ酸ＳとＡの対に対するスコア１を加算して５という値を得、その値が他の場合から算出される１という値よりも大きいので、このマス９０７におけるスコアは５となる。マトリクス９０５の１行目は図７における（２）に対応するマスの値のみを参照して、スコアを算出することになるが、以下、行を重ねるに従って図７に示された７つのマスのスコアを参照して、スコア計算を行うこととなる。この操作を繰り返して、それぞれの

マスに対応するスコアを算出する。マトリクス９０５のマスのスコアを全て計算しものが９０８である。マトリクス９０８の各マス内の円で囲まれた数字がそのマスにおけるスコアであり、左上にある数字は、そのマスにおけるスコアが図７の（１）から（７）までのどのマスのスコアから算出されたかを示す数字である。そして９０８のマトリクス上の１番端の行及び列上において最大値を探しその値を翻訳アミノ酸配列９０３と既知のアミノ酸配列９０４の配列比較におけるスコアとなる。このスコアの大小によって、既知のアミノ酸配列９０４が翻訳アミノ酸配列９０３に類似しているか否かを判断する基準とする。この例の場合では、マトリクス９０８の垂直成分に当たる翻訳アミノ酸配列９０３の最後のアミノ酸Ｔの行と、水平成分に当たる既知のアミノ酸配列９０４の最後のアミノ酸Ｔの列における最大スコアをこの配列比較に対するスコアとする。次に、このスコアの計算結果から翻訳アミノ酸配列９０３と既知のアミノ酸配列９０４の間の並置を求める手順を説明する。並置は、比較が行われた配列間で、配列のどの部分がどのように類似しているかを表示する方法である。並置は最大スコアに対応するマスから、図７の（１）から（７）のどのマスからそのスコアが算出されたかをたどり、（１）から（７）の経路に従って図８のような並置例を繋げていく事によって求められる。この例の場合には、まずマトリクス９０８の最大スコアをとるマス９０９からたどっていくこととなる。マス９０９のスコアは、図７における（２）にあたるマス９１０から計算されているので、９０９から９１０へと並置経路をたどる。マス９１０のスコアも同様に図７の（２）にあたるマスから計算されているので、マス９１１に並置経路をたどる。マス９１１のスコアは、図７の（１）にあたるマス９１２から計算されているので、並置経路は９１１から９１２に飛ぶことになる。このような手順をくり返し、配列を比較した結果の並置経路を求めることが出来る。求められた並置経路に対して、それぞれのマスのスコアが図の（１）から（７）のいずれかのマスのスコアから計算されたかに従って、図８の並置例に従い、並置結果を表示することが出来る。従って、翻訳アミノ酸配列９０３、すなわちＤＮＡ配列９０１と、既知のアミノ酸配列９０４の比較結果としての並置の表示は、９１３に示したようになる。図９で説明に用いた例は、配列が非常に短いために、挿入・欠失に対するペナルティを考慮して計算すると、スコア自身が非常に小さな値になってしまい、検索が出来なくなる。そのため、ここでは配列比較方法の原理を説明するために、挿入・欠失に対するペナルティは考慮しなかったが、実際の検索の時には、〔数１〕に表わされるようなペナルティを導入して、スコア計算を行っている。これはもし、挿入・欠失に対してペナルティを導入しないと、無制限に挿入・欠失をいれてしまうことで、実際には類似していない配列を検索で拾ってきてしまうためである。

【００２５】次に上記で説明された本発明の配列比較方法を用いた検索について述べる。本配列比較方法において、実際に配列をもちいた比較を行なう。以下、アミノ酸配列データベースとして、ＰＩＲ（Ｒｅｌｅａｓｅ３４）の中でｓｕｐｅｒｆａｍｉｌｙ分類の記載のあるデータｐｉｒ１．ｓｅｑ（配列数１０５５０、アミノ酸塩基数３５９１３７０）を利用した。前記既知アミノ酸配列は、アミノ酸配列データベース中に登録されている配列とし、データベース中に含まれているアミノ酸配列のうち実際に翻訳される部分のＤＮＡ配列が分かっているＤＮＡ配列に、配列塩基長の３％にあたる数の塩基の挿入あるいは欠失を生じさせたものをキーＤＮＡ配列として利用した。この値は、実際に解析されたばかりの配列には最悪の場合に３％程の誤りが含まれることを考慮して設定した。従来方法との比較方法としては、元のＤＮＡ配列に対応するアミノ酸配列と類似していると分類されている同じｓｕｐｅｒｆａｍｉｌｙのメンバーをいかに検索で拾ってこれるか否か、または欠失等を生じさせる以前のＤＮＡ配列からの翻訳アミノ酸配列に対して、正しい位置に挿入あるいは欠失を考慮してその並置を求める事が出来るかを評価した。図１０は、従来方法との比較、即ち、どれだけ同じｓｕｐｅｒｆａｍｉｌｙのメンバーを認識できるかを示したグラフである。縦軸に実際に検索で拾ってきたｓｕｐｅｒ　ｆａｍｉｌｙのメンバー数を示しており、横軸には検索で拾ってきたすべての配列数を示している。この結果、配列長の３％の挿入あるいは欠失が存在した場合でも、本発明では従来法よりも効率的に類似したアミノ酸配列を拾ってきている事が分かる。また、図１１には、実際に検索を行った並置結果を示している。１行目が既知アミノ酸配列、２行目にはアミノ酸の一致あるいは不一致等をあらわす記号、３行目には翻訳アミノ酸配列、４行目にはＤＮＡ配列を配置している。各配列中に現れる記号「－」は挿入あるいは欠失がその位置に存在し、該当する塩基が無い事を表わしている。図１１で示されたように、ＤＮＡ配列中に挿入あるいは欠失が存在している場合でも考慮して検索を行っている事が分かる。図１０及び図１１に示された結果より、本発明によるＤＮＡ配列とアミノ酸配列の比較方法は、類似しているアミノ酸配列を従来方法よりも正確に検索できる事が分かる。

【００２６】

【発明の効果】本発明により、新たに決定されたＤＮＡ配列に対して、塩基単位の挿入あるいは欠失を考慮してアミノ酸配列と配列比較を行うことができる。その結果、生体中の機能の解明されているアミノ酸配列に類似しているＤＮＡ配列を見つけ、その類似部分を並置結果として表示することが出来るために、ＤＮＡ配列の持つ機能を類推する事が容易になる。

【００２７】

【配列表】
配列番号：１
配列の長さ：232
配列の型：核酸
鎖の数：一本鎖
トポロジー：直鎖状
配列の種類：mRNA
配列の特徴：ＨＵＭＲＯＳＭＣＦ
トランスメンブラン・プロテインキナーゼの３･末端をコードするヒトmcf3(再配列 ros1)プロト- オンコージンmRNA
配列：
【００２８】

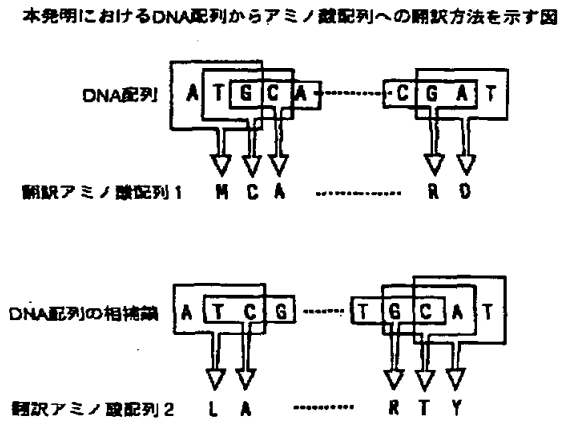
```
AspPheTrpIleProGluThrSerPheIleLeuThrIleIleValGlyIlePheLeuVal
GATTTTTGGATACCAGAAACAAGTTTCATACTTACTATTATAGTTGGAATATTTCTGGTT

ValThrIleProLeuThrPheVal - -  ArgArgLeuLysAsnGlnLysSerAlaLys
GTTACAATCCCACTGACCTTTGTC------TAGAAGATTAAAGAATCAAAAAAGTGCCAAG

GluIleLysValAlaValPro  LysThrLeuLysLysGlySerThrAspGlnGluLysIle
GAAATCAAAGTAGCAGTGCCCC-CAAGACTTTGAAGAAGGGTTCCACAGACCAGGAGAAGATT

LeuThrLeuValAspLeuValAspLeuCysValAspIle - -  GlyCysValTyrLeu
CTCACCTTGGTTGACCTTGTAGACCTGTGTGTAGATATA------AGGCTGTGTCTACTTG
```

【００２９】配列番号：２
配列の長さ：７９
配列の型：アミノ酸
鎖の数：一本鎖
トポロジー：直鎖状配列の種類：ペプチド
配列の特徴：ＴＶＨＵＲＴ　　＃ タイププロテイン(断片)プロテインチロシンキナーゼmcf3(活性型 ros-1) ヒト断片# ＥＣ-No 2.7.1.112
配列：
【００３０】

```
AspPheTrpIleProGluThrSerPheIleLeuThrIleIleValGlyIlePheLeuVal

ValThrIleProLeuThrPheValTrpHis ArgArgLeuLysAsnGlnLysSerAlaLys

GluIleLysValAla  - ValLysThrLeuLysLysGlySerThrAspGlnGluLysIle

LeuThrLeuValAspLeuValAspLeuCysValAspIleSerLys GlyCysValTyrLeu
```

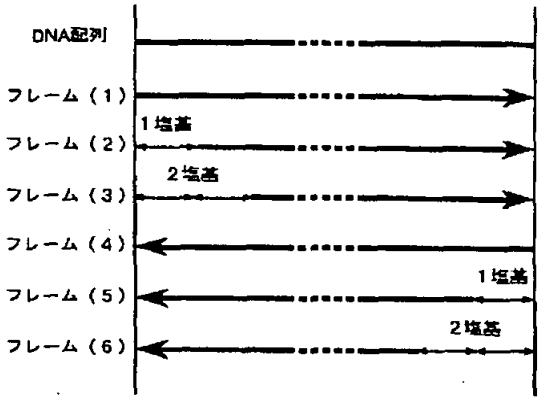
【図面の簡単な説明】
【図１】従来方法におけるＤＮＡ配列からアミノ酸配列への翻訳フレームを示す図。
【図２】Ｓｍｉｔｈ－Ｗａｔｅｒｍａｎ法での配列比較を行なう際のスコアの算出経路を示す図。
【図３】本発明の配列比較方法を適用する配列比較装置の構成を示す図。
【図４】コドンとアミノ酸の対応表を示す図。
【図５】本発明におけるＤＮＡ配列からアミノ酸配列への翻訳方法を示す図。
【図６】アミノ酸同士の対に与えられるスコアの一例を示す図。
【図７】本発明におけるスコア算出時の参照位置を示す図。
【図８】本発明における各スコア参照位置からの経路に対して与えられる並置例を示す図。
【図９】実例を用いた本発明の配列比較の説明図。
【図１０】従来方法との配列比較結果の評価を示す図。
【図１１】本発明における配列比較の並置結果を示す図。

【図５】



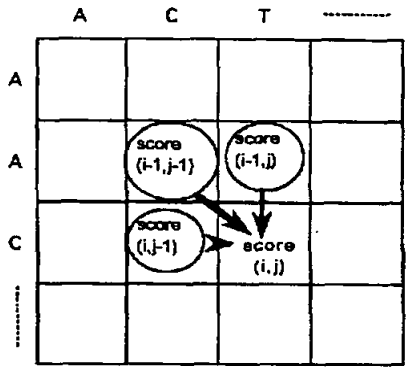
【図１】

従来方法におけるDNA配列からアミノ酸配列への翻訳フレームを示す図



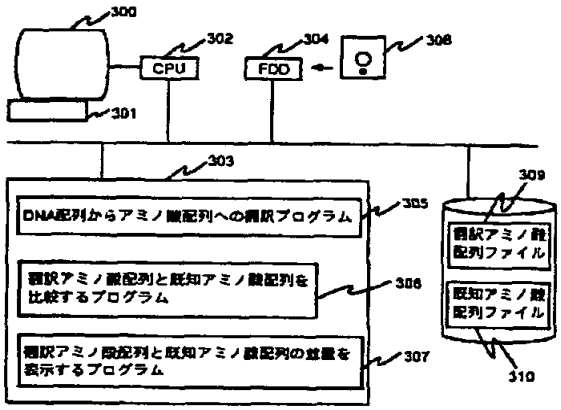
【図２】

Smith-Waterman法での配列比較を行う際のスコアの算出経路を示す図



【図３】

本発明の配列比較方法を適用する配列比較装置の構成を示す図



【図４】

コドンとアミノ酸の対応表を示す図

アミノ酸（三文字表記）	DNA	アミノ酸（三文字表記）	DNA
A (Ala)	GCN	K (Lys)	AAR
R (Arg)	MGN	M (Met)	ATG
N (Asn)	AAY	F (Phe)	TTY
D (Asp)	GAY	P (Pro)	CCN
C (Cys)	TGY	S (Ser)	WSN
Q (Gln)	CAR	T (Thr)	ACN
E (Glu)	GAR	W (Trp)	TGG
G (Gly)	GGN	Y (Tyr)	TAY
H (His)	CAY	V (Val)	GTN
I (Ile)	ATH	終止	TRR
L (Leu)	YTN	X (xxx)	NNN
B (N or D)	RAY	Z (Q or E)	SAR

DNA文字種類対応表

R	A/G	M	A/C
Y	C/T	B	C/G/T
S	C/G	V	A/C/G
W	A/T	D	A/G/T
N	A/C/G/T	H	A/C/T

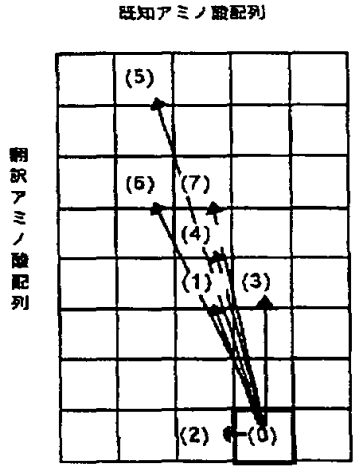
【図６】

アミノ酸同士の対に与えられるスコアの一例を示す図

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0	-2	-1	0	-4
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3	-1	0	-1	-4
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3	3	0	-1	-4
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3	4	1	-1	-4
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1	-3	-3	-2	-4
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2	0	3	-1	-4
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3	-1	-2	-1	-4
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3	0	0	-1	-4
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3	-3	-3	-1	-4
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1	-4	-3	-1	-4
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2	0	1	-1	-4
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1	-3	-1	-1	-4
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1	-3	-3	-1	-4
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2	-2	-1	-2	-4
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2	0	0	0	-4
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0	-1	-1	0	-4
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3	-4	-3	-2	-4
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1	-3	-2	-1	-4
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	-3	-2	-1	-4
B	-2	-1	3	4	-3	0	1	-1	0	-3	-4	0	-3	-3	-2	0	-1	-4	-3	-3	4	1	-1	-4
Z	-1	0	0	1	-3	3	4	-2	0	-3	-3	1	-1	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
X	0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	-2	0	0	-2	-1	-1	-1	-1	-1	-1	-4
*	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	1

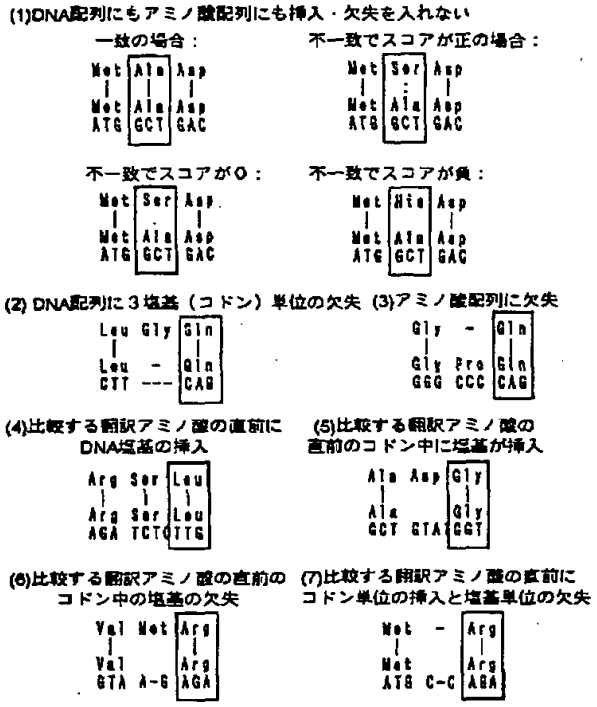
【図７】

本発明におけるスコアの参照位置を示す図



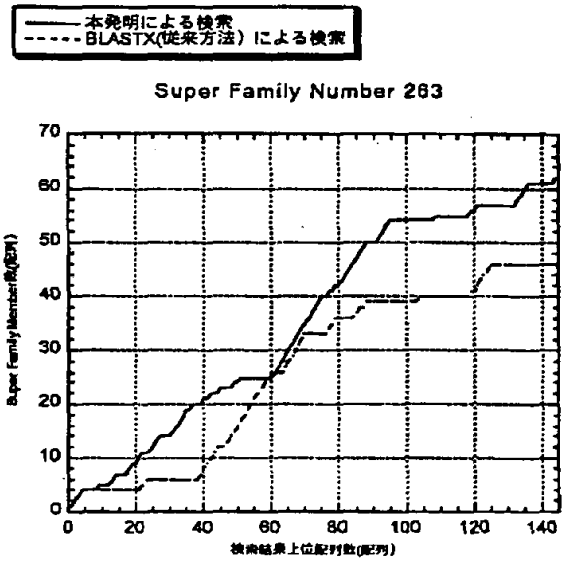
【図８】

各スコア参照位置からの経路に対して与えられる並置例を示す図



【図１０】

従来方法との配列比較結果の評価を示す図

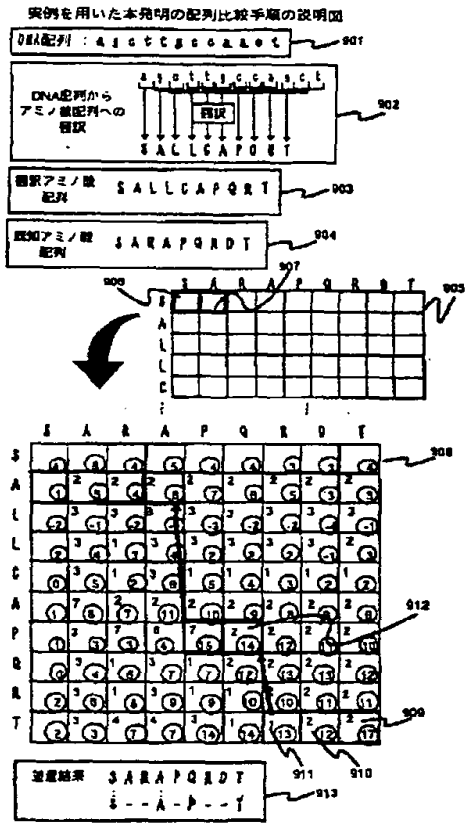


【図１１】

本発明における配列比較の並置結果を示す図

```
Query sequence:>HUMROSMCF (GenBank) Human mcf3 (rearranged ros1)
proto-oncogene mRNA encoding a transmembrane protein kinase, 3' end.

Target sequence:>TVHURT Protein-tyrosine kinase mcf3 (activated ros-1) -
Human (fragment) EC-number 2.7.1.112 | 263.0 22.0 2.0 1.0 2.0

Score: 2333 + strand

Target:  AspPheTrpIleProGluThrSerPheIleLeuThrIleIleValGlyIlePheLeuVal
Query:   AspPheTrpIleProGluThrSerPheIleLeuThrIleIleValGlyIlePheLeuVal
         GATTTTTGGATACCAGAAACAAGTTTCATACTTACTATTATAGTTGGAATATTTCTGGTT

Target:  ValThrIleProLeuThrPheValTrpHis ArgArgLeuLysAsnGlnLysSerAlaLys
Query:   ValThrIleProLeuThrPheVal - -  ArgArgLeuLysAsnGlnLysSerAlaLys
         GTTACAATCCCACTGACCTTTGTC------TAGAAGATTAAAGAATCAAAAAAGTGCCAAG

Target:  GluIleLysValAla -  - ValLysThrLeuLysLysGlySerThrAspGlnGluLysIle
Query:   GluIleLysValAlaValPro  LysThrLeuLysLysGlySerThrAspGlnGluLysIle
         GAAATCAAAGTAGCAGTGCCCC-CAAGACTTTGAAGAAGGGTTCCACAGACCAGGAAGAAGATT

Target:  LeuThrLeuValAspLeuValAspLeuCysValAspIleSerLys GlyCysValTyrLeu
Query:   LeuThrLeuValAspLeuValAspLeuCysValAspIle -  -  GlyCysValTyrLeu
         CTCACCTTGGTTGACCTTGTAGACCTGTGTGTAGATATA------AGGCTGTGTCTACTTG
```

【図９】


実例を用いた本発明の配列比較手順の説明図
DNA配列 : a g c t t g c c a a c t
901
DNA配列から
アミノ酸配列への
翻訳
902
翻訳
S A L L C A P Q R T
翻訳アミノ酸
配列
S A L L C A P Q R T
903
既知アミノ酸
配列
S A R A P Q R D T
904
907
905
906
912
909
911
910
並置結果
S A R A P Q R D T
913